

Chapter 86

Impact of Sarcasm in Sentiment Analysis Methodology

Priscilla Souza Silva

Federal University of South and Southeast of Pará, Brazil

Haroldo Barroso

Federal University of Sul and Sudeste of Pará, Brazil

Leila Weitzel

Fluminense Federal University, Brazil

Dilcielly Almeida Ribeiro

Universidade Federal do Sul e Sudeste do Pará, Brazil

José Santos

Federal University of Sul and Sudeste of Pará, Brazil

ABSTRACT

Sentiment of analysis is a study area applied to numerous environments (financial, political, academic, business, and communication) whose purpose is to search for messages posted on social media, and through these to identify and classify people's opinions about particular item as positive or negative. Rating the sentiment expressed in opinionated messages is such an important task that currently companies invest a lot of money in collecting this type of information and the development of methods and techniques to classify the sentiment that they express, so that they can use the results as useful information in preparing marketing and sales strategies efficiently. However, one of the major problems facing the feelings of analysis is the difficulty of methods to properly analyze messages with sarcastic and/or ironic content, as these linguistic phenomena have the characteristic of transforming the polarity or meaning of a positive or negative statement into its opposite.

DOI: 10.4018/978-1-6684-6303-1.ch086

INTRODUCTION

With the advent of the Web, and increased use of online channels (such as social networks, blogs, social networking sites, online newspapers, forums, recommendations websites and online business tools that allow users to register their opinions on goods and services) in which users exchange information and share their knowledge, criticism, opinions and feelings about a topic of interest, made in the last decade the amount of textual information written in natural language reached gigantic proportions. As Cambria et al. (2013), extract and process properly all this mountain of information has become something extremely interesting to the business world, because through this data companies can get a continuous and faster feedback on the opinion of the public about their products and brand. The academic world in turn is largely responsible for the development of tools and methods to treat such data.

However, despite the relevance of such information, collect and analyze resulting web opinions became an impractical task for the human being in time due to the large amount of textual data published. So to treat and automatically analyze the opinions and sentiments expressed in this type of data has emerged an area called Sentiments Analysis (AS) also called Opinion Mining (LIU, 2010a).

The sentiment analysis according to Benvenuto, Ribeiro and Araújo (2015), is a field of study that uses computer processing to define automated techniques to extract subjective information from texts in natural language, such as opinions and feelings in order to create knowledge structured that can be used by a support system or decision maker. Basically these techniques identify the sentiment that users have regarding any interest entity (a specific product, a company, a place, a person, among others) based on the Web shared texts, allowing a user to get a report containing what people comment on any item, without having to seek and read all the reviews and news about manually.

Currently companies like Walmart, MCDonalds and IBM, are investing in research groups in sentiment analysis, it realized the value of the opinions expressed in social media and how these views may affect them positively or negatively (CHEN, Zimbra, 2010). Other emerging sectors interested in these data are: a policy to keep its members informed about public opinion regarding their actions; famous accompanying its level of popularity among Internet users; producers interested in knowing about the acceptance of the films produced, among other sectors.

Among the different data sources used by the AS, social networks represent the most conducive environment to identify opinions and feelings about different entities, since they are places where people argue about everything expressing political opinions, religious or even about brands, products and services. In addition to these opinions when properly collected and analyzed, allow not only to understand and explain many complex social phenomena, but also provides them.

However, despite the Web and especially social networks facilitating access and distribution of opinionated information, the task of identifying, classifying, and summarizing data views in text format, it is not trivial and has many challenges, which induce inconsistencies the results generated by sentiment analysis applications. Generally, these challenges are related to the difficulty that computers have to automatically process natural language (human). The automation of processing of human language has its principles in the study area called Natural Language Processing (NLP). This area relates directly with AS, they share the study of unstructured data (LIU; HOGAN; CROWLEY, 2011).

Natural language in turn is somewhat complex mainly when it comes to social networking as there are many nuances in a text message. This complexity is a challenge that are factors of human language, such as words and / or phrases with ambiguous meanings, sarcasm and irony, slang, spelling mistakes, regionalism, dialects, among others.

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/impact-of-sarcasm-in-sentiment-analysis-methodology/308566

Related Content

PSSRC: A Web Service Registration Cloud Based on Structured P2P and Semantics

Qian He, Baokang Zhao, Liang Chang, Jinshu Suand Ilsun You (2016). *International Journal of Data Warehousing and Mining* (pp. 21-38).

www.irma-international.org/article/pssrc/146851

U.S. Federal Data Mining Programs in the Context of the War on Terror: The Congress, Courts, and Concerns for Privacy Protections

Shahid M. Shahidullah (2009). *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 151-180).

www.irma-international.org/chapter/federal-data-mining-programs-context/29149

Creating Strategic Information for Organizations with Structured Text

Roberto Penteadand Eric Boutin (2008). *Emerging Technologies of Text Mining: Techniques and Applications* (pp. 34-53).

www.irma-international.org/chapter/creating-strategic-information-organizations-structured/10175

Human-Centred Web Search

Orland Hoeber (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1852-1872).

www.irma-international.org/chapter/human-centred-web-search/73526

An OLAM Operator for Multi-Dimensional Shrink

Stefano Rizzi, Matteo Golfarelliand Simone Graziani (2015). *International Journal of Data Warehousing and Mining* (pp. 68-97).

www.irma-international.org/article/an-olam-operator-for-multi-dimensional-shrink/129525